HW4\_Patrick Neyland

Patrick Neyland

2022-09-13

## Question 1

### Part (i)

Interpret the coefficient on . Is the sign of this estimate what you expect it to be?

This model is an example of one that has undergone a log-log transformation—A 1% increase in distance from the garbage incinerator, holding all else constant, will result in 0.132% increase in price on average. This coefficient is expected—on average, people want to live far away from a presumably stinky and ugly garbage handling facility.

### Part (ii)

What other factors about a house affect its price? Might these be correlated with distance from the incinerator?

Proximity to grocery stores, schools, and rec centers will also affect the price of a home—these specific factors might be correlated with distance from the incinerator. I would imagine that most businesses would also want to be located far away from any garbage handling facilities.

### Part (iii)

Do you think simple regression provides an unbiased estimator of the ceteris paribus elasticity of price with respect to dist? (Think about the city’s decision on where to put the incinerator. Hint: which of the SLR assumptions is/are violated?)

No, it does not provide an unbiased estimator because SLR assumption 4 is violated. The distance from the incinerator is likely correlated with some factor included in the u variable as explained in part ii.

## Question 2

### Part (i)

Show that E[u|inc] = 0, so that the key zero condition mean assumption is satisfied.

Therefore, is equal to .

### Part ii

Show that V ar(u|inc) = einc, so that the homoskedasticity Assumption SLR.5 is violated. In particular, the variance of savings increases with inc. [Hint: Var(e|inc) = Var(e) if e and inc are independent].

### Part (iii)

Provide a discussion that supports the assumption that the variance of savings increases with family income.

As family income increases, the choice of savings amount also increase. Once a family has met its basic needs, it may choose to save any amount it would like. More disposable income will lead to a greater variance in savings because there will be more savings options.

## Question 3

### Part (i)

Find the average participation rate and the average match rate in the sample of plans.

The average participation rate is 87.36 percent. The average match rate is 73.15 percent.

### Part (ii)

Now, estimate the simple regression equation and report the results along with the sample size and R-squared

model3\_2 <- lm(prate~mrate,data = k401k)  
stargazer(model3\_2, type = "text")

===============================================  
 Dependent variable:   
 ---------------------------  
 prate   
-----------------------------------------------  
mrate 5.861\*\*\*   
 (0.527)   
   
Constant 83.075\*\*\*   
 (0.563)   
   
-----------------------------------------------  
Observations 1,534   
R2 0.075   
Adjusted R2 0.074   
Residual Std. Error 16.085 (df = 1532)   
F Statistic 123.685\*\*\* (df = 1; 1532)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

### Part (iii)

Interpret the intercept in your equation. Interpret the coefficient on .

The intercept(83.075) is the is the average participation rate when is zero. The coefficient of shows that on average, a one unit increase in , holding all else constant, will result in a 5.861 percent increase in participation rate.

### Part (iv)

Find the predicted prate when . Is this a reasonable prediction? Explain what is happening here.

prate <- function(mrate) {83.075 + 5.861\*mrate}  
prate(3.5)

[1] 103.5885

The predicted when is 103.59 percent. This is not a reasonable prediction because in reality, cannot exceed 100 percent. The correlation between these two variables is probably not linear.

### Part (v)

How much of the variation in prate is explained by ? Is this a lot in your opinion?

7.5 percent of the variation in is explain by . This does not seem like a lot to me—I suppose it depends.

## Question 4

### Part i

Do you think each additional dollar spent has the same effect on the pass rate, or does a diminishing effect seem more appropriate? Explain.

I think a diminishing effect seems more appropriate. At the beginning each dollar makes a really big difference—buying books, pencils, paper, and basic computers open tremendous opportunities for the students to learn and develop their skills. However, spending an additional $2,000 per student on a fancier computer lab will likely have little impact on test scores because it would not offer significantly more practical functionality than the normal computer lab.

### Part ii

In a normal level-log transformation, needs to be divided by 100 given a 1% change in . Therefore, given a 10% change in would require be divided by only 10. Making this change keeps the scale accurate.

### Part iii

model4\_3 <- lm(math10~log(expend), data = meap93)  
stargazer(model4\_3, type = "text")

===============================================  
 Dependent variable:   
 ---------------------------  
 math10   
-----------------------------------------------  
log(expend) 11.164\*\*\*   
 (3.169)   
   
Constant -69.341\*\*\*   
 (26.530)   
   
-----------------------------------------------  
Observations 408   
R2 0.030   
Adjusted R2 0.027   
Residual Std. Error 10.350 (df = 406)   
F Statistic 12.411\*\*\* (df = 1; 406)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

### Part iv

The estimated spending effect is 1.1164 percent—this can be interpreted as a 10 percent increase in will, on average, result in a 1.1164 percent increase in .

### Part v

Because a level-log transformation was used, decreasing returns are presented in this regression. Therefore is unlikely to pass 100 percent.

## Question 5

### Part i

What is the average gift in the sample of 4,268 people (in Dutch guilders)? What percentage of people gave no gift?

mean(charity$gift)

[1] 7.44447

tb1 <- table(charity$respond)  
ptb1 <- prop.table(tb1)  
ptb1[1]\*100

0   
60.00469

The average gift in the sample is worth 7.44 Dutch guilders.  
The percentage of people who gave no gift is 40 percent.

### Part ii

mean(charity$mailsyear)

[1] 2.049555

min(charity$mailsyear)

[1] 0.25

max(charity$mailsyear)

[1] 3.5

The average number of mailings per year is 2.05.  
The minimum mailings per year is 0.25.  
The maximum mailings per year is 3.5.

### Part iii

model5\_3 <- lm(gift~mailsyear, data = charity)  
stargazer(model5\_3, type = "text")

===============================================  
 Dependent variable:   
 ---------------------------  
 gift   
-----------------------------------------------  
mailsyear 2.650\*\*\*   
 (0.343)   
   
Constant 2.014\*\*\*   
 (0.739)   
   
-----------------------------------------------  
Observations 4,268   
R2 0.014   
Adjusted R2 0.014   
Residual Std. Error 14.960 (df = 4266)   
F Statistic 59.649\*\*\* (df = 1; 4266)   
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

### Part (iv)

Interpret the slope coefficient. If each mailing costs one guilder, is the charityexpected to make a net gain on each mailing? Does this mean the charity makes a net gain on every mailing? Explain.

On average, holding all else constant, if the number of mailings per year increases by 1 unit, the amount of gift will increase by 2.65 guilders. Yes, on average, over the entire data set, the net gain per mailer will be 1.65 guilders. No, the charity will not make a net gain on each mailer—some will be more and some will even be a net loss because the regression just gives the average.

## Question 6

### Part i

count(catholic)

n  
1 7430

mean(catholic$math12)

[1] 52.13362

sd(catholic$math12)

[1] 9.459117

mean(catholic$read12)

[1] 51.7724

sd(catholic$read12)

[1] 9.407761

The sample size 7430 students.  
The average math score is 52.13.  
The standard deviation of math scores is 9.46.  
The average reading score is 51.77.  
The standard deviation of reading scores is 9.41.

### Part ii

model6\_2 <- lm(math12~read12, data = catholic)  
stargazer(model6\_2, type = "text")

===============================================  
 Dependent variable:   
 ---------------------------  
 math12   
-----------------------------------------------  
read12 0.714\*\*\*   
 (0.008)   
   
Constant 15.153\*\*\*   
 (0.432)   
   
-----------------------------------------------  
Observations 7,430   
R2 0.505   
Adjusted R2 0.505   
Residual Std. Error 6.658 (df = 7428)   
F Statistic 7,568.582\*\*\* (df = 1; 7428)  
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

= 15.153  
 = 0.714  
 = 7430  
 = 0.505

### Part iii

It does not have a meaningful interpretation. It is the estimated average math score of students with a score of 0 on the reading test. However, the minimum reading score in the sample was 29.15.

### Part iv

I am not surprised by the , I would have assumed that high reading scores would be an indicator of high math scores—high-performing students like to be high-performing in all subject. I am a little surprised by the value. I figured more the reading score would have been a better predictor of math score.

### Part v

I would say that a correlation between reading and math does not in anyway indicate a causal relationship. I would point him to other factors that are likely to have a much greater impact on math scores such as the number of math tutors or the levels of additional training offered math teachers.